**MEME16203Linear Models**

**Assignment 2**

**UNIVERSITI TUNKU ABDUL RAHMAN**

| | | | |
|---|---|---|---|
| Faculty: | FES | Unit Code: | MEME16203 |
| Course: | MAC | Unit Title: | Linear Models |
| Year: | 1,2 | Lecturer: | Dr Yong Chin Khian |
| Session: | May 2023 | | |
| Due by: | 3/7/2023 | | |

Q1. Show that any two matrices $\mathbf{W}$ and $\mathbf{X}$ have the same column space if there exist matrices $\mathbf{F}$ and $\mathbf{G}$ such that $\mathbf{WG} = \mathbf{X}$ and $\mathbf{XF} = \mathbf{W}$. (10 marks)

Q2. Suppose

$$Y_{ij} = \mu_i + \epsilon_{ij}, i = 1,2; j = 1,2,3.$$

Let $\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23}) \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 36 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 36 & 0 \\ 0 & 0 & 0 & 0 & 0 & 64 \end{bmatrix} \right).$

What is the BLUE of $3\mu_1 + 5\mu_2$? Explain carefully. (15 marks)

Q3. Consider a problem of quartic regression in one variable, $X$. In particular, suppose that $n = 6$ values of a response $y$ are related to values $x = 0, 1, 2, 3, 4, 5$ by a linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$ for

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \\ 1 & 5 & 25 & 125 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \text{ and } \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}$$

Define

$$\mathbf{W} = \begin{bmatrix} 1 & -5 & 5 & -5 \\ 1 & -3 & -1 & 7 \\ 1 & -1 & -4 & 4 \\ 1 & 1 & -4 & -4 \\ 1 & 3 & -1 & -7 \\ 1 & 5 & 5 & 5 \end{bmatrix}$$

(a) Show that $\mathbf{y} = \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$ is reparameterization of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\gamma}^{\boldsymbol{T}} = [\gamma_1, \gamma_2, \gamma_3, \gamma_4]$. (10 marks)

(b) Notice that $\mathbf{W^TW}$ is diagonal. Suppose that $\mathbf{y^T} = (-2, 0, 4, 2, 2, 1)$. Find the OLS estimate of $\boldsymbol{\gamma}$ in the model $\mathbf{y} = \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$ and then OLS estimate of $\boldsymbol{\beta}$ in the original model. (Find numerical values.) (10 marks)

**MEME16203Linear Models**

Q4. Two varieties of corn (variety A and variety B) were compared in a field trail. In addition to the varieties, three levels of nitrogen were used (0, 30, and 60 pounds per acre (lb/a). Six different fields were used, and the six combinations of varieties and nitrogen levels were randomly assigned to the fields. Let $Y_{ij}$ denote the yield (in bushels per acre) of the $i^{th}$ variety of corn when the $j^{th}$ level of nitrogen is applied. Throughout this question, $\epsilon_{ij}$, $i = 1, 2, j = 1, 2, 3$, denote independent $N(0, \sigma^2)$ random variables where $\sigma^2$ is an unknown variance. The following two models were proposed:

$$\text{Model 1:} \quad \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 0 & -30 & 900 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 30 & 900 \\ 0 & 1 & -30 & 900 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 30 & 900 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \end{bmatrix}$$

$$\text{Model 2:} \quad \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & 0 & -2 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & 0 & -2 \\ 1 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \end{bmatrix}$$

(a) With respect to the effects of varieties and nitrogen levels on corn yields, interpret the parameters $\gamma_1$ and $\delta_1$ in Model 1. (5 marks)

(b) For Model 1, indicate which of the following quantities are estimable

$$\gamma_1 - \gamma_2; \qquad \gamma_1 - 10\delta_1 + 100\delta_2$$

Give a brief explanation, to support your conclusions. (5 marks)

(c) For Model 2, determine if $\mu + \alpha_1$ is estimable? Give a brief explanation to support your conclusion. (5 marks)

(d) Expressing Model 2 as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, a solution to the normal equations is $\mathbf{b} = (\mathbf{X^T X})^{-}\mathbf{X^T Y}$. Explain how a generalized inverse $(\mathbf{X^T X})^{-}$ can be computed. (You are not expected to obtain a numerical value for $(\mathbf{X^T X})^{-}$, just briefly outline a procedure for how it can be computed.) (5 marks)

(e) Using $\mathbf{b} = (\mathbf{X^T X})^{-}\mathbf{X^T Y}$ from Part (d), define the estimator

$$\hat{\alpha}_1 - \hat{\alpha}_2 = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \end{bmatrix} \mathbf{b}$$

What are the properties of this estimator? (5 marks)

(f) Would the residual sum of squares from fitting models (1) and (2) be the same? (5 marks)

Q5. Two varieties of corn (variety A and variety B) were compared in a field trail. In addition to the varieties, three levels of nitrogen were used (10, 20 and 30 pounds per acre (lb/a). Six different fields were used, and the six combinations of varieties and nitrogen levels were randomly assigned to the fields. Suppose the data are as follows.

| Field | $i$ | $j$ | Amount of Nitrogen ($x_{ij}$) | Bushels per acre ($y_{ij}$) |
|---|---|---|---|---|
| 1 | 1 | 1 | 10 | 80 |
| 2 | 1 | 2 | 20 | 120 |
| 3 | 1 | 3 | 30 | 140 |
| 4 | 2 | 1 | 10 | 60 |
| 5 | 2 | 2 | 20 | 150 |
| 6 | 2 | 3 | 30 | 170 |

Consider a Gauss-Markov model

$$y_{ij} = \mu + \alpha_i + \gamma_i X_{ij} + \gamma_3 X_{ij}^2 + \epsilon_{ij}$$

where

- $y_{ij}$ is the yield (in bushels per acre) of the $i^{th}$ variety of corn when the $j^{th}$ level of nitrogen is applied.
- $X_{ij}$ denote the level of nitrogen administered to the corn,
- $\mu, \alpha_1, \alpha_2, \gamma_1, \gamma_2, \gamma_3$ are unknown parameters, and
- $\epsilon_{ij}$ denotes a random error with $\epsilon_{ij} \sim NID(0, \sigma^2)$ where $\sigma^2 > 0$.

Let $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \gamma_1, \gamma_2, \gamma_3)^T$, $\mathbf{y} = [y_{11}, y_{12}, y_{13}, y_{21}, y_{22}, y_{23}]^T$, and $\boldsymbol{\epsilon} = [\epsilon_{11}, \epsilon_{12}, \epsilon_{13}, \epsilon_{21}, \epsilon_{22}, \epsilon_{23}]^T$.

(a) Determine the design matrix $\mathbf{X}$ so that $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$. (5 marks)

(b) Determine whether $\alpha_2$ is estimable. Prove that your answer is correct. (5 marks)

(c) Show that $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{W})$, where

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & -1 & 0 & 1 \\ 1 & 1 & 0 & 0 & -2 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & -1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 0 & -2 \\ 1 & -1 & 0 & 1 & 1 \end{bmatrix}.$$

(5 marks)

(d) Verify that $\tau = \mu + \alpha_1$ is estimable, then obtained the unique BLUE of $\tau$. (10 marks)